

Collect Data

type of variable	categorical vs quantitative (continuous vs discrete)
type of descriptive methods	tabular, graphic, Numerical
Tabular	n, f, rf, 100rf, cf, rcf, 100rcf
graphical	relations: bar, pie, dot, stem leave, histogram, cumulative freq
numeric	precise/inference, dull, complicated

graph the data

qualitative	bar pie
quantitative	dot plot, stemplot; histogram, cumulative freq charts, boxplot
Examining graph	center(mean, median, mode), spread (range, std, variance), shape (symmetric, skewed)
pattern/deviations	cluster/gap, outliers
dotplot	spread, shape, approx center
stemplot	shaped, spread, center
histograph	f vs rf, shape/center, large dataset, error bar for spread
cumulative freq charts	S shaped, T(skewed) shape, meaningful order

Central tendency - mean, median, mode

summering distribution	population/sample, center/spread/sape
mean	μ =population mean, \bar{x} bar=sample mean

Central tendency - mean, median, mode (cont)

median	for skewed data, odd/even sample size
mode	number with highest freq
symmetrical	mean=median=mode
left skewed	mode>median>mean
right skewed	mode>mean>media

variance/spread - range, IQR, STD

variance	spread from mean
range	largest-smallest measurement, outliers affect
IQR	interquartile range, eg Q3-Q1, not affected by outliers, median / IQR
STD	standard deviation, square root of variance, outlier affect, ≥ 0
variance	average the square of deviation from mean
population variance	N, sigma, μ
sample variance	n-1, \bar{x} bar, s
mean/STD, median/IQR	

Position - quartile, percentile, standardized score

percentile	order, divide into 100 equal parts, count kth percentile P_k
quartile	order, divide into 4 equal parts (median calc), count kth quartile Q_k , $P_{25}=Q_1$, $P_{50}=Q_2$.
z score	standardized score, $(x-\text{mean})/\text{std}$, compare datasets with different scales, eg temperature in north vs south city

Graphing uni variant data

graphical summaries	Y scale: misleading manipulation
box plots	box(Q2-Q3) and whiskers(-lower Q1, upper Q4), whiskers $< 1.5\text{IQR}$ (Q_3-Q_1), $L=Q_1-1.5\text{IQR}$, $U=Q_3+1.5\text{IQR}$. point $>U$ or $<L$ are outliers
	based on position, identify outlier and general shape(-skewed or not)
	calc: Q1, Median, Q3, IQR, L, U
shift unit +a	(variance/spread) range, std, IQR not affected
enlarge or shrink unit, *b	all stat enlarged or shrunk
Compare distributions	center, spread, shape
	outlier or unusual values, cluster/gap
	context of the question
	dot plot, stemplot, histogram, freq polygram

Avoid simple list the stat (center, std and shape), instead, make a clear comparative statement.

Bivariant data

Scatter plot	shape: linear, non-linear, no relation
	direction: positive or negative linear relation
	strength of linear relation: close to the line
Numeric methods	
correlation coefficient	degree and direction of linear relation of two quantitative variables (x,y)
	ρ and r, [-1, +1]



Bivariant data (cont)

0, 0.1, 0.5, 0.85, 1

least squares regression line

formular $Y = a + bX + e$

Y dependent/response variable

x independent/explanatory variable

a y intercept of line

b slope of the line

e random error, residual error

predicted y hat
vaue

residual e
error

least square regression minimize the sum of squares of residual error

line of best fit (\bar{X} , \bar{Y}),
slope = $r(S_y/S_x)$

coefficient of determination R squared, percent of variance of Y determined by variance X

[-1, +1]

influential point point that affect the correlation efficient

Outlier maybe influential point

residual plot should be random, or else, fit is not the best

transformation to fit linear log, sqrt, reciprocal, square, power

1 calc slope, intercept, write formula, plot the linear line

2 make a prediction, calc residual error

3 calc coefficient of determination $r = \frac{SS_{xy}}{\sqrt{SS_{xx} * SS_{yy}}}$

stat and interpretation

categorical data

marginal and joint freq of two way tables

contingency(- r*c joint) table

marginal row col grand total

conditional relative frequency

association compare with row total * col total / grand total



By **Jianmin Feng** (taotao)
cheatography.com/taotao/

Not published yet.

Last updated 17th December, 2019.

Page 2 of 2.

Sponsored by **CrosswordCheats.com**

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>